

Testing convexity of a discrete distribution

Fadoua Balabdaoui¹, Cécile Durot^{2*} and François Koladjo³

¹CEREMADE, Université Paris-Dauphine, 75775, Paris, France

²Modal'x, Université Paris Nanterre, F-92001, Nanterre, France

³ENSPD, Université Parakou, BP 55 Tchaourou, Bénin

Abstract

Based on the convex least-squares estimator, we propose two different procedures for testing convexity of a probability mass function supported on \mathbb{N} with an unknown finite support. The procedures are shown to be asymptotically calibrated.

1 The testing problem

Estimating a probability mass function (pmf) under a shape constraint has attracted attention in the very last years, see Jankowski and Wellner (2009), Durot et al. (2013), Balabdaoui et al. (2013), Giguélay (2016). With a more applied point of view, Durot et al. (2015) developed a nonparametric method for estimating the number of species under the assumption that the abundance distribution is convex. As the method applies only if the convexity assumption is fulfilled, it would be sensible, before to implement it, to test whether or not the assumption is fulfilled. This motivates the present paper where a method for testing convexity of a pmf on \mathbb{N} is developed.

We consider i.i.d. observations X_1, \dots, X_n from an unknown pmf p_0 on \mathbb{N} . Assuming that p_0 has a finite support $\{0, \dots, S\}$ with an unknown integer $S > 0$, we aim at testing the null hypothesis H_0 : " p_0 is convex on \mathbb{N} " versus the alternative H_1 : " p_0 is not convex." With p_n the empirical pmf (defined by $p_n(j) = n^{-1} \sum_{i=1}^n \mathbb{1}_{\{X_i=j\}}$ for all $j \in \mathbb{N}$), a natural procedure rejects H_0 if p_n is too far from \mathcal{C}_1 , the set of all convex probability mass functions on \mathbb{N} . Hence, with $\|q\|^2 = \sum_{j \in \mathbb{N}} (q(j))^2$ for a sequence $q = \{q(j), j \in \mathbb{N}\}$, we reject H_0 if $\inf_{p \in \mathcal{C}_1} \|p_n - p\|^2$ is too large. It is proved in Durot et al. (2013, Sections 2.1 to 2.3) that the minimizer exists, is unique, and can be implemented with an appropriate algorithm, so our critical region takes the form $\{T_n > t_{\alpha,n}\}$ where $T_n = \sqrt{n} \|p_n - \hat{p}_n\|$, \hat{p}_n is the minimizer of $\|p_n - p\|^2$ over $p \in \mathcal{C}_1$, and $t_{\alpha,n}$ is an appropriate quantile. The main difficulty now is to find $t_{\alpha,n}$ in such a way that the corresponding test has asymptotic level α .

We consider below two different constructions of $t_{\alpha,n}$. First, we will define $t_{\alpha,n}$ to be the $(1 - \alpha)$ -quantile of a random variable whose limiting distribution coincides with the

*Corresponding author. Email address: cecile.durot@gmail.com

limiting distribution of T_n under H_0 . Next, we will calibrate the test under the least favorable hypothesis. Both methods require knowledge of the limiting distribution of T_n under H_0 . To this end, we need notation. For all $p = \{p(j), j \in \mathbb{N}\}$ and $k \in \mathbb{N} \setminus \{0\}$ we set $\Delta p(k) = p(k+1) - 2p(k) + p(k-1)$ (hence p is convex on \mathbb{N} iff $\Delta p(k) \geq 0$ for all k) and a given $k \in \mathbb{N} \setminus \{0\}$ is called a knot of p if $\Delta p(k) > 0$. Furthermore, we denote by g_0 a $(S+2)$ centered Gaussian vector whose dispersion matrix Γ_0 has component $(i+1, j+1)$ equal to $\mathbb{1}_{\{i=j\}}p_0(i) - p_0(i)p_0(j)$ for all $i, j = 0, \dots, S+1$, and by \hat{g}_0 the minimizer of $\sum_{k=0}^{S+1} (g(k) - g_0(k))^2$ over the set \mathcal{K}_0 of all functions $g = (g(0), \dots, g(S+1)) \in \mathbb{R}^{S+2}$ such that $\Delta g(k) \geq 0$ for all $k \in \{1, \dots, S\}$ with possible exceptions at the knots of p_0 . Existence, uniqueness and characterization of \hat{g}_0 are given in Balabdaoui et al. (2014, Theorem 3.1). The asymptotic distribution of T_n under H_0 is given below.

Theorem 1.1. *Under H_0 , $T_n \xrightarrow{d} \hat{T}_0$ as $n \rightarrow \infty$, where $\hat{T}_0 = \sum_{k=0}^{S+1} (\hat{g}_0(k) - g_0(k))^2$.*

2 Calibrating by estimating the limiting distribution

In order to approximate the distribution of T_n under H_0 , we will construct a random variable that weakly converges to \hat{T}_0 (see Theorem 1.1) and which can be approximated via Monte-Carlo simulations. To this end, let $S_n = \max\{X_1, \dots, X_n\}$. Also, let g_n be a random vector which, conditionally on (X_1, \dots, X_n) , is distributed as a centered Gaussian vector of dimension $S_n + 2$ with dispersion matrix Γ_n , the matrix with component $(i+1, j+1)$ equal to $\mathbb{1}_{\{i=j\}}p_n(i) - p_n(i)p_n(j)$ for all $i, j = 0, \dots, S_n+1$. Now, let \hat{g}_n be the minimizer of $\sum_{k=0}^{S_n+1} (g(k) - g_n(k))^2$ over a set \mathcal{K}_n that approaches \mathcal{K}_0 as $n \rightarrow \infty$. Below, we give an extended version of Balabdaoui et al. (2014, Theorem 3.3), with the same choice for \mathcal{K}_n .

Theorem 2.1. *Let $(v_n)_{n \in \mathbb{N}}$ be a sequence of positive numbers that satisfy $v_n = o(1)$ and $v_n \gg n^{-1/2}$. Define g_n and \hat{g}_n as above with \mathcal{K}_n the set of all functions $g = (g(0), \dots, g(S_n+1)) \in \mathbb{R}^{S_n+2}$ such that $\Delta g(x) \geq 0$ for all $x \in \{1, \dots, S_n\}$ that satisfy $\Delta \hat{p}_n(x) \leq v_n$. Then, \hat{g}_n uniquely exists, both \hat{g}_n and $\hat{T}_n := \sum_{k=0}^{S_n+1} (\hat{g}_n(k) - g_n(k))^2$ are measurable, and conditionally on X_1, \dots, X_n we have $\hat{T}_n \xrightarrow{d} \hat{T}_0$ in probability as $n \rightarrow \infty$, with \hat{T}_0 as in Theorem 1.1.*

We now state the main result of the section, again defining \hat{T}_n as in Theorem 2.1.

Theorem 2.2. *Let $\alpha \in (0, 1)$ and $t_{\alpha, n}$ the conditional $(1-\alpha)$ -quantile of \hat{T}_n given X_1, \dots, X_n . If p_0 is convex on \mathbb{N} and supported on $\{0, \dots, S\}$, then $\limsup_{n \rightarrow \infty} P(\hat{T}_n > t_{\alpha, n}) \leq \alpha$.*

In order to implement the test, we need to compute an approximation of $t_{\alpha, n}$. This can be done using Monte-Carlo simulations as follows. Having observed X_1, \dots, X_n , draw independent sequences $(Z_i^{(b)})_{1 \leq i \leq S_n}$ for $b \in \{1, \dots, B\}$, where all variables $Z_i^{(b)}$ are i.i.d. standard Gaussian and $B > 0$ is an integer. Then, for all b , compute $g_n^{(b)} = \Gamma_n^{1/2}(Z_0^{(b)}, \dots, Z_{S_n+1}^{(b)})^T$ and the least-squares projection $\hat{g}_n^{(b)}$ onto \mathcal{K}_n using the algorithm described in Balabdaoui et al. (2014). Then, $t_{\alpha, n}$ can be approximated by the $(1-\alpha)$ -quantile of the empirical distribution corresponding to $\sum_{k=0}^{S_n+1} (g_n^{(b)}(k) - \hat{g}_n^{(b)}(k))^2$, with $b \in \{1, \dots, B\}$.

3 Calibrating under the least favorable hypothesis

We consider below an alternative calibration that is easier to implement than the first one. Consider $\tilde{\mathcal{K}}_0$ the set of convex functions g on $\{0, \dots, S+1\}$; that is $g \in \tilde{\mathcal{K}}_0$ if and only if $\Delta g(k) \geq 0$ for all $k \in \{1, \dots, S\}$. Similarly, let $\tilde{\mathcal{K}}_n$ be the set of convex functions on $\{0, \dots, S_n+1\}$. Let \tilde{g}_0 be the least squares projection of g_0 onto $\tilde{\mathcal{K}}_0$ and \tilde{g}_n that of g_n onto $\tilde{\mathcal{K}}_n$, with g_0 and g_n as in Section 2. Finally, let $\tilde{t}_{n,\alpha}$ be the conditional $(1-\alpha)$ -quantile of $\tilde{T}_n := \sum_{k=0}^{S_n+1} (\tilde{g}_n(k) - g_n(k))^2$ given (X_1, \dots, X_n) . Then, we have the following theorem.

Theorem 3.1. *If p_0 is convex on \mathbb{N} and supported on $\{0, \dots, S\}$, then $\limsup_{n \rightarrow \infty} P(T_n > \tilde{t}_{n,\alpha}) \leq \alpha$ with equality if p_0 is the triangular pmf with support $\{0, \dots, S\}$.*

The test is asymptotically calibrated since the Type I error does not exceed α . It reaches precisely α when p_0 is triangular, which can be viewed as the least favorable case for testing convexity. The theorem above does not exclude existence of other least favorable cases.

4 Simulations

To illustrate the theory, we have considered four pmf's supported on $\{0, \dots, 5\}$. It follows from Theorem 7 in Durot et al. (2013) that any convex pmf on \mathbb{N} can be written as $\sum_{k \geq 1} \pi_k \mathcal{T}_k$ where $\pi_k \in [0, 1]$, $\sum_{k \geq 1} \pi_k = 1$ and $\mathcal{T}_k(i) = 2(k-i)_+[k(k+1)]^{-1}$, the triangular pmf supported on $\{0, \dots, k-1\}$. Under H_0 , we considered the triangular pmf $p_0^{(1)} = \mathcal{T}_6$ and $p_0^{(2)} = \sum_{k=1}^6 \pi_k \mathcal{T}_k$ with $\pi_1 = 0, \pi_2 = \pi_3 = 1/6, \pi_4 = 0$ and $\pi_5 = \pi_6 = 1/3$, which has knots at 2, 3 and 5. Under H_1 we considered $p_1^{(1)}$ the pmf of a truncated Poisson on $\{0, \dots, 5\}$ with rate $\lambda = 1.5$ and $p_1^{(2)}$ the pmf equal to $p_0^{(1)}$ on $\{2, \dots, 5\}$ such that $(p_1^{(2)}(0), p_1^{(2)}(1)) = (p_0^{(1)}(0) + 0.008, p_0^{(1)}(1) - 0.008)$. To investigate the asymptotic type I error and power of our tests, we have drawn $n \in \{500, 5000, 50000\}$ rv's from the aforementioned pmf's. Here, $\alpha = 5\%$ and $\tilde{t}_{n,\alpha}$ was estimated for each drawn sample using $B = 1000$ i.i.d. copies of g_n . The rejection probability was estimated using $N = 500$ replications of the whole procedure. For the first convexity test, we considered the sequences $v_n \equiv \sqrt{\log(\log n)n^{-1/2}}$ and $n^{-1/4}$. We also added the sequence $v_n \equiv 0$ to compare our approach with the naive one where no knot extraction is attempted. The results are reported in Tables 1 and 2.

The first conclusion is that choosing $v_n = 0$ does not give a valid test as the type I error can be as large as four times the targeted level! This can be explained by the fact that choosing $v_n = 0$ makes the set on which g_n is projected to be the largest possible and hence $\|\hat{g}_n - g_n\|$ the smallest possible. This distance is hence stochastically smaller than the actual limit, yielding a large probability of rejection. The second conclusion is that the first test depends on the choice of v_n . Small sequences makes again the type I error large when the true convex pmf has only a tiny change in the slopes at its knots or has no knots as it is the case for $p_0^{(1)} = \mathcal{T}_6$. The question is then open as to how to choose such a

PMF	$n = 500$			$n = 5000$			$n = 50000$		
	0	$\frac{(\log(\log n))^{1/2}}{n^{1/2}}$	$n^{-1/4}$	0	$\frac{(\log(\log n))^{1/2}}{n^{1/2}}$	$n^{-1/4}$	0	$\frac{(\log(\log n))^{1/2}}{n^{1/2}}$	$n^{-1/4}$
$p_0^{(1)}$	0.226	0.106	0.054	0.286	0.092	0.062	0.256	0.086	0.050
$p_0^{(2)}$	0.190	0.046	0.020	0.310	0.066	0.018	0.344	0.054	0.016
$p_1^{(1)}$	1	1	1	1	1	1	1	1	1
$p_1^{(2)}$	0.234	0.102	0.038	0.354	0.166	0.082	0.932	0.816	0.630

Table 1: Values of the asymptotic type I error for the pmfs $p_0^{(1)}$ and $p_0^{(2)}$ and power for $p_1^{(1)}$ and $p_1^{(2)}$ of the convexity test based on the sequence $(v_n)_n$. The asymptotic level is 5%.

PMF	$n = 500$	$n = 5000$	$n = 50000$
$p_0^{(1)}$	0.048	0.044	0.058
$p_0^{(2)}$	0.014	0.032	0.020
$p_1^{(1)}$	1	1	1
$p_1^{(2)}$	0.042	0.060	0.678

Table 2: Values of the asymptotic type I error for the pmfs $p_0^{(1)}$ and $p_0^{(2)}$ and power for $p_1^{(1)}$ and $p_1^{(2)}$ of the test based on the least favorable hypothesis. The asymptotic level is 5%.

sequence so that the test has the correct asymptotic level. The second testing approach is, as expected, conservative when the true pmf is not triangular. For the true pmf $p_1^{(1)}$ which strongly violates the convexity constraint, the power is equal to 1. For $p_1^{(2)}$, which has only a small flaw at 2 with a change of slope equal to -0.008 , the power values obtained with this second approach are comparable to those obtained with the first testing method and $v_n \equiv n^{-1/4}$. This is somehow expected as it is the largest sequence among the ones considered, yielding the largest distance between g_n and its L_2 projection.

5 Proofs

In the sequel, for all $s > 0$ and $u = (u(0), \dots, u(s+1)) \in \mathbb{R}^{s+2}$, we set $\|u\|_s = \sum_{k=0}^{s+1} (u(k))^2$.

5.1 Preparatory lemmas

Lemma 5.1. *Let $s > 0$ be an integer, $\mathcal{K} \subset \mathbb{R}^{s+2}$ a non-empty closed convex set, and $u = (u(0), \dots, u(s+1)) \in \mathbb{R}^{s+2}$. Then, the minimizer of $\|g - u\|_s$ over $g \in \mathcal{K}$ uniquely exists. Moreover, denoting $\Phi(u)$ this minimizer, the application Φ is measurable from \mathbb{R}^{s+2} to \mathbb{R}^{s+2} ; and the application $u \mapsto \|u - \Phi(u)\|_s$ is measurable.*

Proof: It follows from standard results on convex optimization that $\Phi(u)$ uniquely exists

for all u , and $\|\Phi(u) - \Phi(v)\|_s \leq \|u - v\|_s$ for all u and v in \mathbb{R}^{s+2} . This means that Φ is a continuous function, whence it is measurable. Now, the function $u \mapsto (u, \Phi(u))$ is continuous, whence measurable. By continuity of the norm, this ensures that the application that maps u into $\|u - \Phi(u)\|_s$ is measurable. \square

Lemma 5.2. *With g_0 as in Section 1, $(\Delta g_0(1), \dots, \Delta g_0(S))$ is a centered Gaussian vector with invertible dispersion matrix.*

Proof: For notational convenience, we assume in the sequel that $S \geq 3$. The case $S \leq 2$ can be handled likewise. Let B be the $S \times (S+1)$ -matrix which j -th line has components $j, j+1$ and $j+2$ equal to 1, -2 and 1 respectively while the other components are zero, for $j = 1, \dots, S-1$, and S -th line has components equal to zero except the penultimate and the last one, which are equal respectively to 1 and -2 . We have $p_n(S+1) = p_0(S+1) = 0$ almost surely so that in the limit, $g_0(S+1) = 0$ almost surely and

$$(\Delta g_0(1), \dots, \Delta g_0(S))^T = B (g_0(0), \dots, g_0(S))^T. \quad (5.1)$$

Hence, $(\Delta g_0(1), \dots, \Delta g_0(S))$ is a centered Gaussian vector with dispersion matrix

$$V = B \Sigma_0 B^T, \quad (5.2)$$

where Σ_0 is the dispersion matrix of the vector on the right hand side of (5.1), *i.e.* with component $(i+1, j+1)$ equal to $\mathbb{1}_{\{i=j\}} p_0(i) - p_0(i)p_0(j)$ for all $i, j = 0, \dots, S$. Note that Σ_0 is obtained by deleting a line and a column of zeros in Γ_0 , the dispersion matrix of g_0 .

It remains to prove that V is invertible. Let $\sqrt{p_0}$ be the column vector in \mathbb{R}^{S+1} with components $\sqrt{p_0(0)}, \dots, \sqrt{p_0(S)}$ and let $\text{diag}(\sqrt{p_0})$ be the $(S+1) \times (S+1)$ diagonal matrix with diagonal components $\sqrt{p_0(0)}, \dots, \sqrt{p_0(S)}$. Denoting by I the identity matrix on \mathbb{R}^{S+1} , the matrix (in the canonical basis) associated with the orthogonal projection from \mathbb{R}^{S+1} onto the orthogonal supplement of the linear space generated by $\sqrt{p_0}$ is given by $I - \Pi_0 = I - \sqrt{p_0} \sqrt{p_0}^T$. The linear subspace of \mathbb{R}^{S+1} generated by $\sqrt{p_0}$ has dimension 1, so its orthogonal supplement in \mathbb{R}^{S+1} has dimension S , whence $\text{rank}(I - \Pi_0) = S$. Now, $\Sigma_0 = \text{diag}(\sqrt{p_0})(I - \Pi_0)\text{diag}(\sqrt{p_0})$ where $\text{diag}(\sqrt{p_0})$ is invertible, so we obtain

$$\text{rank}(\Sigma_0^{1/2}) = \text{rank}(\Sigma_0) = S. \quad (5.3)$$

This means that the kernel of $\Sigma_0^{1/2}$ is a linear subspace of \mathbb{R}^{S+1} whose dimension is equal to 1. Let us describe more precisely the kernel. Let λ be the column vector in \mathbb{R}^{S+1} whose components are all equal to 1. Using that $\sum_{k=0}^S p_0(k) = 1$, it is easy to see that $\Sigma_0 \lambda$ is the null vector in \mathbb{R}^{S+1} . Therefore, $\lambda^T \Sigma_0 \lambda = 0$. This means that $\|\Sigma_0^{1/2} \lambda\|^2 = 0$ with $\|\cdot\|$ the euclidian norm in \mathbb{R}^{S+1} . Hence, $\Sigma_0^{1/2} \lambda$ is the null vector in \mathbb{R}^{S+1} . This means that the kernel of $\Sigma_0^{1/2}$ is the linear subspace of \mathbb{R}^{S+1} generated by λ .

Next, let us determine the kernel of V . Let $\mu \in \mathbb{R}^S$ with $V\mu = 0$. Then, $\mu^T V \mu = 0$ which, according to (5.2), implies that $\|\Sigma_0^{1/2} B^T \mu\|^2 = \mu^T B \Sigma_0 B^T \mu = 0$. This means that

$\Sigma_0^{1/2} B^T \mu = 0$. Since the kernel of $\Sigma_0^{1/2}$ is the linear subspace of \mathbb{R}^{S+1} generated by λ , we conclude that $B^T \mu = a\lambda$ for some $a \in \mathbb{R}$. Denote by μ_1, \dots, μ_S the components of μ . By definition of B and λ , the vector μ satisfies the equations $\mu_1 = a$, $\mu_2 - 2\mu_1 = a$, $\mu_{k-2} - 2\mu_{k-1} + \mu_k = a$ for all $k \in \{3, \dots, S\}$ and $\mu_{S-1} - 2\mu_S = a$. Arguing by induction, we obtain that this is equivalent to $2\mu_k = ak(k+1)$ for all $k \in \{1, \dots, S\}$ and $2\mu_S = \mu_{S-1} - a$. Combining the first equation with $k = S, S-1$ to the second equation yields

$$aS(S+1) = -a + \frac{a(S-1)S}{2}.$$

Therefore,

$$a \left(1 - \frac{(S-1)S}{2} + S(S+1) \right) = 0.$$

This reduces to $a(2 + S^2 + 3S)/2 = 0$, which implies that $a = 0$. This means that μ is the null vector in \mathbb{R}^S and therefore, $\text{rank}(V) = S$. This means that V is invertible. \square

Lemma 5.3. *Let \mathcal{K}_0 , g_0 and \widehat{g}_0 be defined as in Section 1. Then, $\|\widehat{g}_0 - g_0\|_S$ has a continuous distribution.*

Proof: Let F be the cumulative distribution function of $\|\widehat{g}_0 - g_0\|_S$. Note that this quantity is a properly defined random variable by the measurability proved in Lemma 5.1. We aim to prove that F is a continuous function on $[0, \infty)$, using similar arguments as in the proof of Lemma 1.2 in Gaenssler et al. (2007). First, we will prove that

$$F(t) > 0 \quad \text{for all } t \geq 0. \tag{5.4}$$

To this end, note that $F(0) = \mathbb{P}(g_0 \in \mathcal{K}_0)$. This means that

$$F(0) \geq \mathbb{P}(\delta \in A) \tag{5.5}$$

with $\delta = (\Delta g_0(1), \dots, \Delta g_0(S))$ and A the set of all vectors $(u_1, \dots, u_S) \in \mathbb{R}^S$ such that $u_k \geq 0$ for all $k \in \{1, \dots, S\}$ with possible exceptions at points k that are knots of p_0 . From Lemma 5.2, the vector δ is a centered Gaussian vector whose dispersion matrix is invertible. Therefore, the vector possesses a density with respect to the Lebesgue measure on \mathbb{R}^S that is strictly positive on the whole space \mathbb{R}^S . This implies that the probability that the vector belongs to a Borel set whose Lebesgue measure is not equal to zero, is strictly positive. In particular, $\mathbb{P}(\delta \in A) > 0$. Thus, it follows from (5.5) that $F(0) > 0$. Combining this with the monotonicity of the function F completes the proof of (5.4).

Next, we prove that the function $\log(F)$ (which is well defined on $[0, \infty)$ thanks to (5.4)) is concave on $[0, \infty)$. For $u = (u(0), \dots, u(S+1)) \in \mathbb{R}^{S+2}$, let us write \widehat{u} the minimizer of $\sum_{k=0}^{S+1} (g(k) - u(k))^2$ over $g \in \mathcal{K}_0$. For all $t \in [0, \infty)$, we define A_t to be the set of all $u = (u(0), \dots, u(S+1)) \in \mathbb{R}^{S+2}$ such that $\|\widehat{u} - u\|_S \leq t$. Note that A_t is a Borel set in $\mathcal{B}(\mathbb{R}^{S+2})$ for all t since, according to Lemma 5.1, the application $u \mapsto \|\widehat{u} - u\|_S$ is

measurable. Finally, let $\mu = \mathbb{P} \circ g_0^{-1}$ be the distribution of g_0 on \mathbb{R}^{S+2} endowed with the Borel σ -algebra $\mathcal{B}(\mathbb{R}^{S+2})$. This means that

$$F(t) = \mu(A_t). \quad (5.6)$$

Fix $\lambda \in (0, 1)$, $t, t' \in [0, \infty)$, and consider an arbitrary $x \in \lambda A_t + (1 - \lambda)A_{t'}$. Then, x takes the form $x = \lambda u + (1 - \lambda)v$ for some (non necessarily unique) $u \in A_t$ and $v \in A_{t'}$. By definition, both \hat{u} and \hat{v} belong to the convex set \mathcal{K}_0 and therefore, $\lambda \hat{u} + (1 - \lambda)\hat{v} \in \mathcal{K}_0$. Since \hat{x} minimizes $\|g - x\|_S$ over $g \in \mathcal{K}_0$, we conclude that

$$\|\hat{x} - x\|_S \leq \|\lambda \hat{u} + (1 - \lambda)\hat{v} - x\|_S \leq \|\lambda(\hat{u} - u) + (1 - \lambda)(\hat{v} - v)\|_S,$$

using that $x = \lambda u + (1 - \lambda)v$. It then follows from the triangle inequality that

$$\|\hat{x} - x\|_S \leq \lambda \|\hat{u} - u\|_S + (1 - \lambda)\|\hat{v} - v\|_S.$$

Since $u \in A_t$ and $v \in A_{t'}$, we have $\|\hat{u} - u\|_S \leq t$ and $\|\hat{v} - v\|_S \leq t'$, which implies that $\|\hat{x} - x\|_S \leq \lambda t + (1 - \lambda)t'$. Hence, $x \in A_{\lambda t + (1 - \lambda)t'}$. This means that

$$\lambda A_t + (1 - \lambda)A_{t'} \subset A_{\lambda t + (1 - \lambda)t'}. \quad (5.7)$$

Now, $\mu = \mathbb{P} \circ g_0^{-1}$ is a Gaussian probability measure on $\mathcal{B}(\mathbb{R}^{S+2})$, so it follows from Lemma 1.1 in Gaenssler et al. (2007) that μ is log-concave in the sense that

$$\mu_*(\lambda A + (1 - \lambda)B) \geq \mu(A)^\lambda \mu(B)^{1 - \lambda}$$

for all $\lambda \in (0, 1)$ and $A, B \in \mathcal{B}(\mathbb{R}^{S+2})$, with μ_* the inner measure pertaining to μ . Applying this with $A = A_t$ and $B = A_{t'}$, and combining with (5.7) yields

$$\mu_*(A_{\lambda t + (1 - \lambda)t'}) \geq \mu(A_t)^\lambda \mu(A_{t'})^{1 - \lambda}$$

for all $t, t' \in [0, \infty)$. The same inequality remains true with μ_* replaced by μ . Using (5.6), and taking the logarithm on both sides of the inequality, we conclude that

$$\log(F(\lambda t + (1 - \lambda)t')) \geq \lambda \log(F(t)) + (1 - \lambda) \log(F(t'))$$

for all $\lambda \in (0, 1)$, and $t, t' \in [0, \infty)$. This means that the function $\log(F)$ is concave on $[0, \infty)$. Recalling (5.4), we conclude that the function $\log(F)$ is continuous on $[0, \infty)$, whence F is continuous on $[0, \infty)$. This completes the proof of Lemma 5.3. \square

5.2 Proofs of the main results

Proof of Theorem 1.1: Assume that p_0 is convex on \mathbb{N} and supported on $\{0, \dots, S\}$. It can be proved in the same manner as Theorem 3.2 in Balabdaoui et al. (2014) that

$$\sqrt{n}(\hat{p}_n - p_0, p_n - p_0) \Rightarrow (\hat{g}_0, g_0) \text{ as } n \rightarrow \infty. \quad (5.8)$$

as a joint weak convergence on $\{0, \dots, S+1\}$. Now, it follows from Balabdaoui et al. (2014, Proposition 3.5) that with probability one, \hat{p}_n is supported on $\{0, \dots, S+1\}$ for sufficiently large n , and \hat{p}_n also is supported on that set by definition. Hence, $\|p_n - \hat{p}_n\| = \|p_n - \hat{p}_n\|_S$. Combining this with (5.8) completes the proof of the theorem. \square

Proof of Theorem 2.1: Clearly, \mathcal{K}_n is a non-empty closed convex subset of \mathbb{R}^{S_n+2} . Hence, we consider the specific case of $s = S_n$ and $\mathcal{K} = \mathcal{K}_n$ in Lemma 5.1. In the notation of the lemma, we have $\hat{g}_n := \Phi(g_n)$, so that \hat{g}_n is uniquely defined. Moreover, since both Φ and g_n are measurable, we conclude that $\hat{g}_n = \Phi(g_n)$ is measurable. Likewise, $\|g_n - \hat{g}_n\|_{S_n}$ is measurable. This proves the first two assertions in Theorem 2.1. Next, similar to Balabdaoui et al. (2014, Theorem 3.3), the following joint weak convergence on $\{0, \dots, S+1\}$ can be proved: conditionally on X_1, \dots, X_n , $(\hat{g}_n, g_n) \Rightarrow (\hat{g}_0, g_0)$ in probability as $n \rightarrow \infty$. The result follows, since $S_n = S$ with probability that tends to one. \square

Proof of Theorem 2.2: Assume that p_0 is convex on \mathbb{N} with support $\{0, \dots, S\}$. By Theorem 1.1, T_n converges in distribution to $\|\hat{g}_0 - g_0\|_S$ as $n \rightarrow \infty$. Combining this with Lemma 5.3 together with the fact that convergence in distribution to a continuous distribution implies uniform convergence of the corresponding distribution functions yields

$$P(T_n > t_{\alpha,n}) = P(\|\hat{g}_0 - g_0\|_S > t_{\alpha,n}) + o(1).$$

Now, it follows from Theorem 2.1 that

$$P(\|\hat{g}_0 - g_0\|_S > t_{\alpha,n}) = P(\|\hat{g}_n - g_n\|_{S_n} > t_{\alpha,n} \mid X_1, \dots, X_n) + o_p(1)$$

where by definition of $t_{\alpha,n}$, the probability on the right-hand side is less than or equal to α for all n . Combining this with the preceding display completes the proof. \square

Proof of Theorem 3.1: We have $\|\tilde{g}_0 - g_0\|_S \geq \|\hat{g}_0 - g_0\|_S$ since $\mathcal{K}_n \subset \mathcal{K}_0$, whence

$$\begin{aligned} P(T_n > \tilde{t}_{\alpha,n}) &= P(\|\hat{g}_0 - g_0\|_S > \tilde{t}_{\alpha,n}) + o(1) \\ &\leq P(\|\tilde{g}_0 - g_0\|_S > \tilde{t}_{\alpha,n}) + o(1), \end{aligned} \tag{5.9}$$

using similar arguments as for the proof of Theorem 2.2 for the equality. Using arguments similar to those in Theorem 3.3 of Balabdaoui et al. (2014), we can show that conditionally on X_1, \dots, X_n , $(\tilde{g}_n, g_n) \rightarrow (\tilde{g}_0, g_0)$ almost surely as $n \rightarrow \infty$. It follows that conditionally on X_1, \dots, X_n , $\|\tilde{g}_n - g_n\|_{S_n}$ converges weakly to $\|\tilde{g}_0 - g_0\|_S$ almost surely whence

$$\begin{aligned} P(T_n > \tilde{t}_{\alpha,n}) &\leq P(\|\tilde{g}_n - g_n\|_{S_n} > \tilde{t}_{\alpha,n} \mid X_1, \dots, X_n) + o(1) \\ &= \alpha + o(1), \text{ by definition of } \tilde{t}_{\alpha,n}. \end{aligned}$$

The inequality in (5.9) becomes an equality if p_0 is triangular, so the theorem follows. \square

References

BALABDAOUI, F., DUROT, C. and KOLADJO, F. C. (2014). On asymptotics of the discrete convex lse of a pmf. *To appear in Bernoulli*.

- BALABDAOUI, F., JANKOWSKI, H., RUFIBACH, K. and PAVLIDES, M. (2013). Asymptotic distribution of the discrete log-concave mle and some applications. *JRSS-B*. **75** 769–790.
- DUROT, C., HUET, S., KOLADJO, F. and ROBIN, S. (2013). Least-squares estimation of a convex discrete distribution. *Comput. Statist. Data Anal.* **67** 282–298.
URL <http://dx.doi.org/10.1016/j.csda.2013.04.019>
- DUROT, C., HUET, S., KOLADJO, F. and ROBIN, S. (2015). Nonparametric species richness estimation under convexity constraint. *Environmetrics* **26** 502–513.
- GAENSSLER, P., MOLNÁR, P. and ROST, D. (2007). On continuity and strict increase of the cdf for the sup-functional of a gaussian process with applications to statistics. *Results in Mathematics* **51** 51–60.
- GIGUELAY, J. (2016). Estimation of a discrete probability under constraint of k-monotony. *arXiv preprint arXiv:1608.06541*.
- JANKOWSKI, H. K. and WELLNER, J. A. (2009). Estimation of a discrete monotone distribution. *Electronic journal of statistics* **3** 1567.